Virtual Workload Measurement for Assessing Systems Utilizing Automation Technology

Matthew A. C. Guckenberger*, Alicia M. Sudol†, and Dimitri N. Mavris‡
Georgia Institute of Technology, Atlanta, Georgia, 30332

With the increasing integration of automation technologies, the role of the operator is changing from sole actor to a shared supervisor/actor role. Studies on unmanned ground vehicle operators and recent crashes partially blamed on automation technologies demonstrate the need to measure and assess operator awareness and workload. Overcoming these challenges requires an assessment early in the design cycle for operator awareness and workload. This methodology integrates concepts from cognitive engineering into operations analysis to better capture and analyze the effectiveness of increasingly automated systems. An agent-based model is created using Operational Event Sequence Diagrams and concepts from situation awareness research to guide agent formulation. The agent rule set is then mapped to the NASA Task Load Index scales to provide a dynamic output throughout the simulation. A traffic model is built in AFSIM to compare the mental workload associated with city versus highway driving. The dynamic workload measurement is the first step in a framework which will enable automation technologies to be traded during the conceptual design phase.

I. Introduction

Automation presents many benefits to system effectiveness; however, integration comes with unique challenges. This work discusses a methodology for assessing new automation technologies. Over the past decade, automation has increasingly made its way into system design, with the eventual goal being system autonomy. A 2018 ANSYS White Paper emphasized the importance that autonomy will play in the future across all industries, "Autonomous vehicles are threatening to disrupt the automotive, aerospace and industrial equipment industries with the emergence of self-driving cars, drones and mobile autonomous robots. They promise to drastically reduce accidents, minimize congestion, bring mobility to the immobile and perform mundane or hazardous tasks in a fraction of the time required by human-controlled vehicles" [1]. However, this progression is coming in stages and it will be a substantial amount of time before it can be fully realized. This can be best understood through the system automation levels.

The Society of Automotive Engineers (SAE) has agreed on descriptors for different automation levels; these levels provide a method through which technologies can be discussed. Figure 1 shows the five levels of automation. Level 1 represents a system which is primarily controlled by the operator while level 5 represents an autonomous system capable of handling any situation. Levels 2 and 3 represent the transition between the operator or the technology functioning as the primary decision maker. The advancements within the automotive industry demonstrate these stages. Manufacturers (e.g. Tesla and Cadillac) and service providers (e.g. Uber and Waymo) are all seeking new automation technologies to give them a competitive edge [3]. This has created automation features such as automatic emergency braking, lane keeping, and more recently mostly automatic driving; however, none of the current solutions enable full autonomy, whereby the technology is capable of independently adapting to any situation that it encounters [3]. Similar technologies are being fielded within the aerospace and defense industries to improve operational effectiveness. One such system is the Boeing Airpower Teaming System which pairs unmanned aircraft with manned aircraft to increase capabilities while decreasing vulnerability; it is expected to first fly in 2020 [4]. The Teaming System allows countries to achieve increased airpower without requiring expansive air force resources (pilots and primary aircraft). The system is not currently meant to achieve autonomy, rather provide additional capabilities to the primary pilot through automation. The pilot would direct the actions of the unmanned aircraft while the unmanned aircraft increase the perception and armament capabilities of a singular aircraft. This teaming between the pilot and the technology results in shared responsibilities, similar to those experienced in partially-automated vehicles, and presents unique operational challenges which are not expected to dissipate soon. Even once full autonomy is achieved in a system, it will take a substantial amount of time to

*Senior Graduate Researcher, Daniel Guggenheim School of Aerospace Engineering, Student Member
†Research Engineer II, Daniel Guggenheim School of Aerospace Engineering, Member
‡S.P. Langley NIA Distinguished Regents Professor, Daniel Guggenheim School of Aerospace Engineering, Fellow AIAA

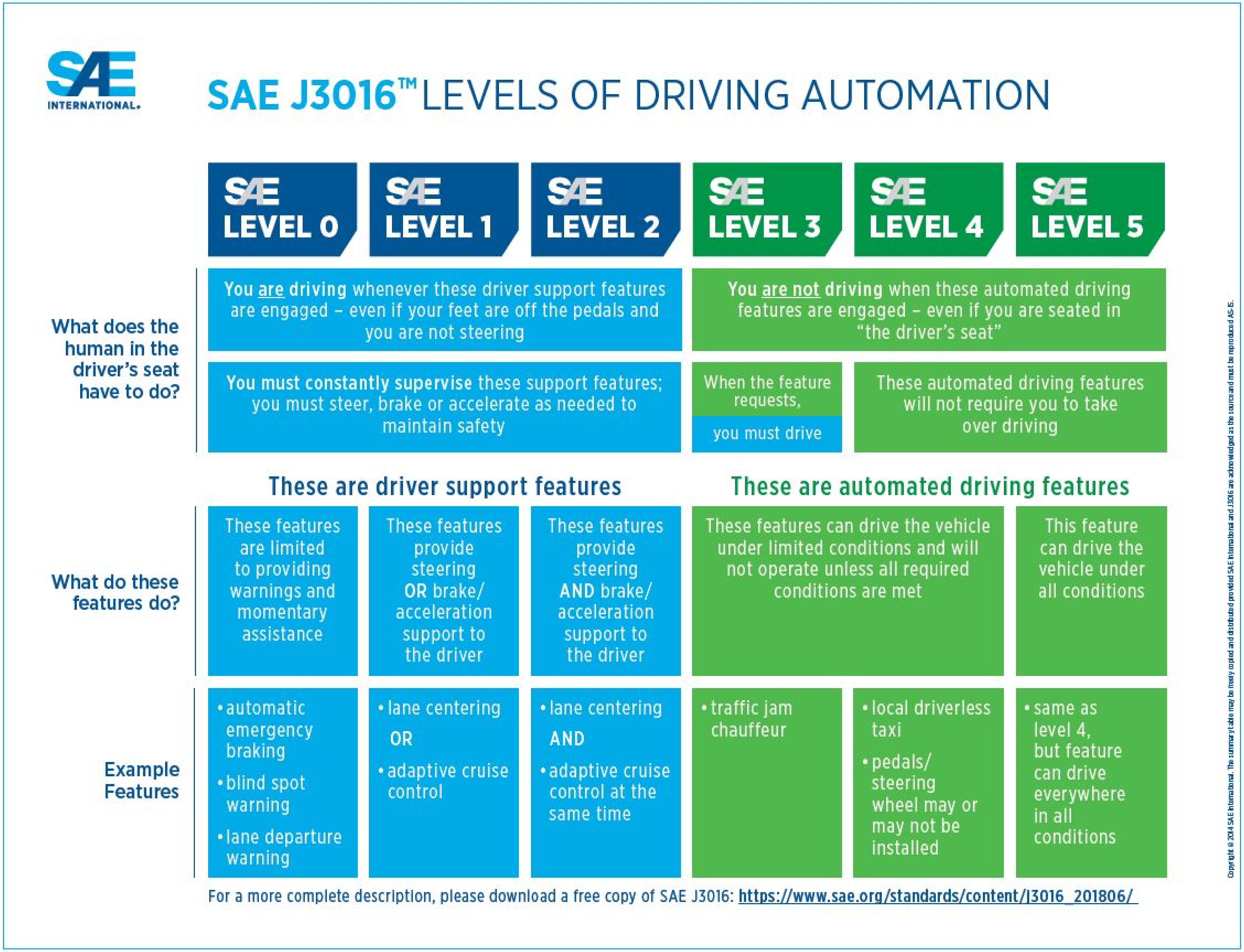

Fig. 1 SAE Standard for Levels of Driving Automation [2].

test the system to ensure appropriate failure rates [5]. As technology slowly progresses towards autonomy and while the testing is conducted, the operator remains an integral component in system utilization. However, their role is changing as the level of automation effects the teaming between the operator and the technology.

This work focuses on the first step of assessing automation technologies, the creation of a virtual, quantitative operator workload metric based on explicit tasking. A methodology is discussed that utilizes cognitive engineering principles to create improved operation analysis studies capable of capturing the operator's workload throughout a scenario. The operator's actions are modeled using operational event sequence diagrams with decision points based on the operator's awareness level, not the systems. This differentiation between system and operator awareness is in line with the work done by Endsley [6–8] regarding situation awareness. The detailed actions are mapped to a quantitative workload metric based on the NASA Task Load Index. This methodology is tested utilizing a virtual driving study, where the operator workload is assessed in an agent-based model and compared with the findings of a similar study conducted with volunteers. While this paper focuses on the creation and validation of a metric, the methodology can be expanded to assess automation technologies early in the system design process. Automation technologies change the form in which tasks are conducted and the operator's awareness of different environmental elements. Therefore, a proper technology assessment requires a reduction in uncertainty around this automation and operator team.

II. Background

During the transition from no-automation to autonomy (from SAE level 0 to level 5), the operator and technology function as a team. *Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity* discusses joint activity among people and the four basic requirements involved: 1) an agreement to work together, 2) mutually predictable in actions, 3) mutually directable, and 4) maintain a common ground [9]. The work by Kirschenbaum [10] discusses the ten challenges experienced when these same requirements are applied to automation integration. Requirement 1 and challenge 1 focus on the concept of the basic compact. The basic compact represents the understanding between the operator and technology to work together, and the resulting limitations of the teamwork, "breakdowns occur when a party abandons the team without clearly signaling his or her intentions to others" [9]. To accurately portray the operational effectiveness of an automation system, it is important to understand which functions the operator controls and which fall to the technology. Similarly, challenge 9 highlights the impact that teamwork has on attention management. Both teammates must have the information necessary for the decisions that they are required to make. In simulating the effectiveness of automation it is important that the operator is modeled in a manner consistent with the attention they would likely have in a partially automated team and that the workload is captured throughout a system's operational cycle. When the effect of automation integration is not understood with respect to workload, awareness, and performance, the improper technology could be chosen thereby reducing performance or worse could result in unexpected system failures. To overcome these challenges, systems utilizing automation technologies must be assessed as a team (operator and technology). Analyses must account for awareness and workload for each team member separately to better capture system performance and identify possible automation issues. This would allow designers to understand when awareness is low due to a low operator workload, such as in the case with the Tesla and the 737 Max.

The integration of automation technologies (SAE levels 1-3) is creating an unmeasured amount of uncertainty within early system design and virtual testing. The autonomy paradox states that "the very systems designed to reduce the need for human operators require more manpower to support them" [11]. Although authors state it differently, the concept is always the same. These technologies change the role of the operator, with functions now shared among the team and new functions required (e.g. technology supervision), sometimes causing an undesirable amount of workload or awareness to the operator or operators. Although the system is capable of perceiving and/or performing properly, the operator or technology may be unable to do so due to a lack of awareness or inappropriate amount of workload. This is clearly indicated by reviewing the recent accidents relating to the Tesla Autopilot and 737 MAX Autopilot. A 2019 NTSB report after a Tesla crashed into a parked firetruck found that "the probable cause for the crash was the Tesla driver's lack of response to the fire truck parked in his lane, due to his inattention and overreliance on the car's advanced driver assistance system" [12]. This represents a breakdown in understanding between the driver and automation resulting in an inappropriate amount of awareness and system failure. The system requires no input/workload from the operator, thus the driver is allowed to completely disengage even though he remains responsible for collision mitigation. A similar automation issue is seen by the crashes of the Beoing's 737 Max. Although design flaws in the Maneuvering Characteristics Augmentation System (MCAS) was the primary factor in the accidents, it has also brought up concerns around pilot skills and intervention capabilities. Automated systems are flipped on for ninety percent of a typical commercial trip which creates an automation dependence [13]. A veteran pilot, John Cox, stated that "automation dependence is not a cause, but it is a contributor' to the disaster" [13], and similar remarks were stated in Transportation Department Inspector General Calvin Scovel's prepared remarks for the US Senate subcommittee on Aviation and Space, "Boeing 737 MAX 8 accidents have suggested a possible link to one of the aircraft's automation systems, raising concerns about pilots' abilities to recognize and react to unexpected events" [14]. The increasing level of system automation and reliance is creating a need to better understand the role of the operator. A lack of insight into which system-level functions are handled by the operator and technology, relating to both workload and awareness, provides an incomplete picture required for evaluations. To overcome the human-agent challenges identified above, a reduction in uncertainty between the role of the operator and technology is required.

The effects of operator workload and awareness on performance are increasingly being studied. The Army Research Lab (ARL) conducted studies in 2007 [15] and 2009 [16] which combined the role of gunner and robotics operator in a military multi-tasking environment to understand the effect of automation on system performance. The gunner role was responsible for local target identification, while an unmanned ground vehicle (UGV) sought out remote targets. Operators were responsible for target identification both locally and remote with identification success monitored. Each trial tested operator identification success while changing the level of UGV automation. Some of the results from the 2009 study are shown in Figure 2. It shows that there is a negative correlation between gunnery task performance and workload. This study also showed best UGV performance when the operator was utilizing teleoperated control (most involved in the task). Through this investigation, ARL showed the impact of changing automation levels on performance and workload based on teaming responsibilities. The system performance is directly dependent on the implementation and control methodology, and the level of automation must be assessed in context to provide an understanding of operator workload and awareness levels, or system performance could suffer. Based on a performance threshold and desired workload level, a proper automation technology and control methodology may be chosen within the system design process. However, this requires an operations model capable of assessing these factors. The need for a more detailed model for assessing automation is also seen in a study looking at air traffic controllers.

Conducted in 2016, the study looked at the time taken to detect conflicts for ATCs utilizing different levels of automation technology [17]. The results of this study are shown in Figure 3. The ATC always had to identify the conflicts (through the entering of a key code when detected), but they could also be responsible for routine tasks (e.g. passing aircraft in and out of airspace) and decision making (e.g. conflict resolution). The perceived workload was higher for ATCs with higher tasking, as expected; however, of interest is that the performance was not a linear negative correlation with workload. Instead performance peaked when the ATC was still responsible for routine tasking. When the ATC had routine tasking, they were required to maintain a satisfactory level of awareness in alignment with the task of conflict detection without being overburdened by a high level of workload. This concept is in alignment with the Hebb/Yerkes-Dodson hybrid, which relates arousal to performance [18]. Hebb/Yerkes-Dodson hybrid showed that performance has a bell-shaped relationship with arousal. There is a desired amount of workload that produces the best performance. Although this bell-shaped relationship has been considered an over simplification of the relationship between performance and workload, it does highlight the importance of assessing the operator's workload and awareness levels when assessing new automation technologies.

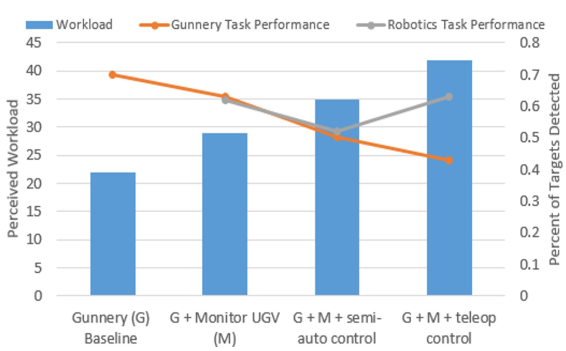

Fig. 2 Army Research Lab (ARL) study results showing target detection percentages for a gunner with different levels of UGV control [16].

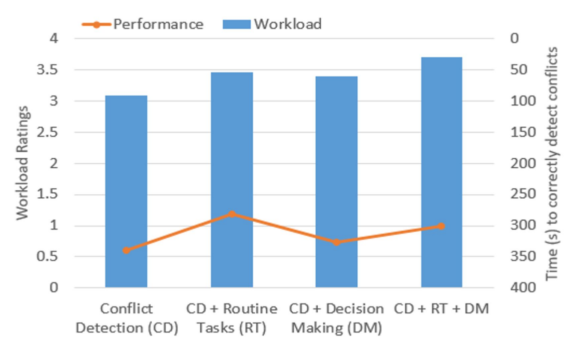

Fig. 3 Study results showing air traffic controllers' (ATC) conflict detection times based on different levels of automation [17].

These volunteer based studies provide validity to the creation of a virtual workload metric but are too costly, time-consuming, and difficult to be the sole provider of operator information early in the system design process.

The operations model must be capable of providing similar insights for systems including automation. The field of operations research provides the contextual studies required to assess system performance, however, care must be taken in modeling selection and model creation to enable the insight into operator awareness and workload instead of system-level awareness and task load. This operator focus is satisfied by including elements from cognitive analysis. Figure 4 shows the relative relationship between the three engineering fields. Operations research and system design are already combined to provide performance trades, while cognitive analysis and system design are combined to better understand the operator's interactions with a system through Hierarchical Task Analysis (HTA). However, these human factors studies are often carried out late in the design process, when many decisions have already been made and the system is considerably constrained [19]. By combining all three fields, virtual studies can be conducted which are capable of capturing the teaming between a technology and the operator in a dynamic environment to create a quantitative workload metric and assess automation technologies. The notional connection of the pieces required is shown in Figure 5. The dynamic task list represents a detailed action set that the operator follows based on his perception of the environment, driven by the sensory model. The workload model represents the quantitative output based on the operator's tasking. Each component requires information from cognitive sciences to guide the model creation.

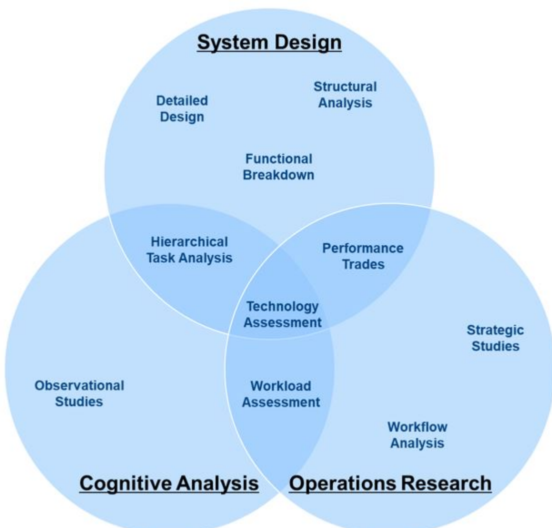

Fig. 4 Notional overlap between systems design, operations research, and cognitive analysis required to assess automation technologies.

Fig. 5 The three pieces that must be added to the operations model to reduce uncertainty around operator workload.

The first component is the dynamic task list. This list is a behavioral breakdown composed based on the system's goal or goals. Operational Event Sequence Diagrams (OESDs) can be utilized to visualize the actions of a system. OESDs are constructed based on the action to be completed and the who or what that will perform each task [20]. The OESD shows the information-decision-action sequence and can be utilized to create detailed action sets which are connected to perception elements. A detailed example OESD for a car changing lanes can be seen in "A Comprehensive Examination of Naturalistic Lane-Changes" [21]. OESDs are executed based on the intent of the driver (e.g. subject vehicle (SV) decides to change lanes which is the result of a different action/goal). Based on this desire, information-action-decision steps are followed, determining which mirrors are checked (SA improvements) and actions are taken. These rules are modified with the introduction of automation (e.g. certain scans may be lower priority if also checked by technology). Once the task list is defined, it is important to map these actions to the dependencies in awareness.

The sensory model must have an operator focus to ensure that the teaming effect can be captured. Current operations models focus on the system-level, if it can be perceived then it is utilized. However, this gives an inaccurate representation of system performance, because each team member (the automation and operator) is aware of different elements. Endsley's research has focused on the concept of situation awareness (SA), and broke it down to the framework and three levels shown in Figure 6 [6] [7] [8]. Endsley's model discusses the three levels of awareness: perception, comprehension, and projection. These levels are important to understand when determining how the task list should be followed. Operators have varying levels of information and require time to achieve SA (e.g. mirror glance times range from 0.8

to 1.6 sec) [21]. Similarly shown by the Endsley's research is the effect that the operator's goals and preconceptions have on his SA. An operator will acquire SA that is inline with his current goal. Utilizing information from SA studies, such as knowledge acquisition times, and the model on Endsley's SA framework will allow the operator to be captured independent from the system. This change allows the operator's tasks to be driven by his perception and the resulting actions to be measured.

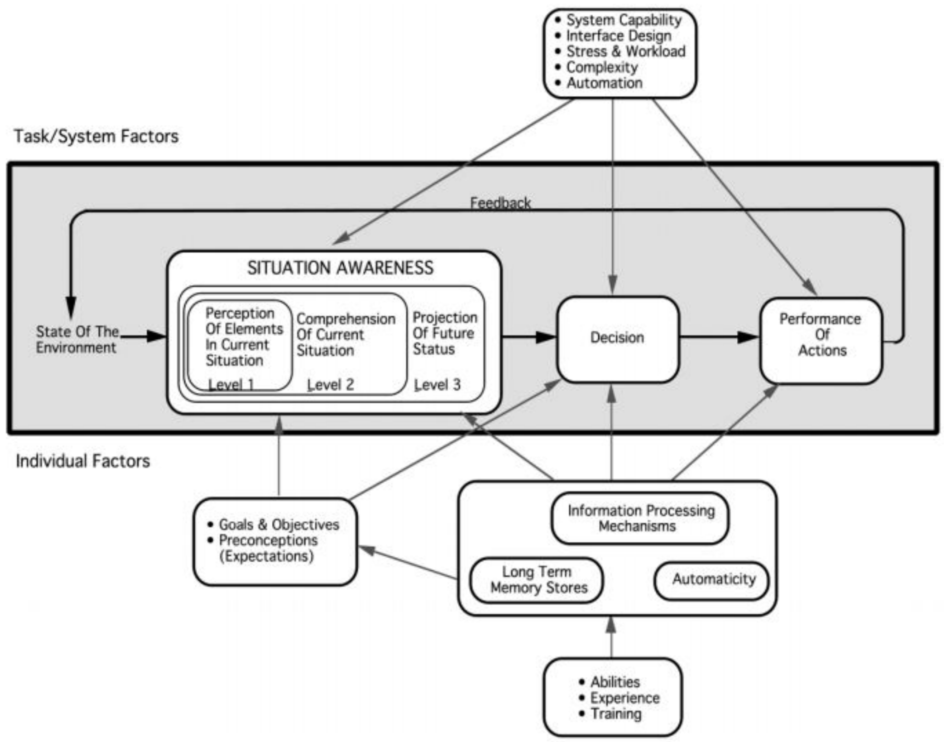

Fig. 6 Endsley model for situation awareness [7].

The final component is the workload model which is used to quantify the operator's demands. This study derives its measurement methodology from the NASA Task Load Index (TLX). TLX is one of the most commonly referenced workload measurement tools [22] and is utilized to categorize the operator tasks. The TLX scale uses a questionnaire given to operators/participants to gauge their workload among six scales [23]. It is one of the most prevalent workload measurement techniques due to its general applicability and simplicity. This framework will focus on mapping tasks to the mental and temporal demands, although this preliminary step only looks at mental demand. Mental demand is composed of the number of tasks currently being handled by the operator, while temporal demand will increase when actions are required before full SA can be achieved. Performance will also be tracked but from the system-level (actual success) not the operator's perceived success. This study is currently only looking at navigation tasks, but as different types of tasks are conducted by the operator they must be mapped together. It is intended that the framework will use data from physiological studies, such as eye tracking, to map lower-level stress to quantitative numbers which can then be binned based on workload type (mental or temporal). For example, Martin et. al [24] used eye-tracking to study the mental workload of air traffic controllers performing tasks with and without conflicts. These variations in workload can be mapped to the actions and events in the operations model to capture the resultant workloads throughout the simulation.

III. Approach

In order to overcome some of the uncertainty of automation technologies and assess there adequacy for integration, it is important that system performance during operations analysis be a function of operator awareness and workload. Similarly, it is important that the workload be measured dynamically throughout the analysis to identify operational locations where the workload may be locally higher or lower than desired. Creating a model with a dynamic task list, operator perception component, and workload measurement provides a virtual, cognition metric for assessing system's utilizing automation technologies. Agent-based modeling is utilized because of the similarities between autonomous systems and the structure of software agents [25]. Agent-based models capture system effects through a bottom-up simulation approach, where agents are intelligent objects that act based on their independent rule sets [25] [26]. Modifying the rule set changes the behaviors of the agents and thus the resulting performance of the system of systems. This low-level definition is helpful in defining which actions (physical and awareness) are taken by the operator versus by the technology. It also enables an iterative development approach for the framework.

The test case looks at a driving scenario due to the prevalence of data. Situation awareness and action-based operator studies are available to create the scenario while results from volunteer-based workload studies are available for comparison. The agent action can be created based on OESDs such as the one created by the NHTSA [21]. Where OESDs are unavailable in published literature then one is created based on knowledge of the driving activity. The perception activities of the operator are based on the OESD and current goals, while the operator's knowledge set changes in accordance with SA studies (e.g. knowledge is dropped when no longer required by the operators goal set). Perception activities will occur based on time-based dependencies. The operator's (driver) actions are tracked throughout the scenario and compiled to provide a dynamic workload metric. For simplicity and time, the first focus is on mental demand, such as perception activities, while temporal demand, such as time-forced decision making, and performance, based on the time to arrive at his destination and the number of errors along the trip (speeding, near misses, emergency braking) will be saved for the further development.

This agent-based model is implemented in the Advanced Framework for Simulation, Integration and Modeling (AFSIM). AFSIM is becoming the Department of Defense's primary tool for up to mission-level analysis [27, 28]. It was chosen due to its open-architecture and class based nature. The open-architecture ensured maximum flexibility in agent behavior creation, while the classes allow quick agent and environment creation. One primary type of agent was created, a car with a single driver. The agent has a rule set governed by the OESD, but each agent is allowed to independently proceed through the environment. Environments are specified based on the layout of the roads, density of traffic, and governing conditions (ex: speed). Cars are randomly placed inside the scenario with a randomly assigned destination and must work to traverse to that destination. Route planning is based on the shortest path to their destination.

Changes in modeling technique through the inclusion of cognitive engineering better portrays operator behaviors and provides a quantitative workload metric. Measurement is analyzed at the agent level to understand the variations in workloads experienced. An agent or set of agents is chosen and the workloads are captured. The mean workloads are then qualitatively compared to a volunteer-based study done by Rahman, Dawal, & Yusoff [29]. Since TLX uses a self-reported scale, it is difficult to get the same results, however, trends between different environments should remain similar. The virtual approach also provides peak and valley data for each workload, unlike volunteer-based studies. Since this paper focuses solely on quantification of workload, the peaks and valleys cannot be compared. However, as this methodology is expanded, multiple runs with different technologies included will change the duration and amplitude of peaks and valleys, as discussed by Schneider, McGrogan, Colombi, Miller, & Long [30]. These quantifications provide necessary metrics for the evaluation of automation technologies in early system design.

Through the utilization of cognitive engineering concepts (OESD, SA and TLX) and combining them with systems design and operations research principles, a model can be created capable of dynamic workload measurement. This methodology ensures that the system performance accounts for the teaming effect and enables the workload to be analyzed for undesirable peaks or valleys. By improving the virtual analysis to separate the operator from the system, uncertainty surrounding automation technology introduction is reduced and trades may be conducted early in the design process in a similar manner to tradition system parameters.

IV. Preliminary Results

The model and complete test cases are still under development, but the framework's early data is showing promise. Online data from OpenStreetMap [31] has been utilized to create a downtown and highway road network within AFSIM. The data for the networks are shown in Figure 7 and Figure 8 respectively. Atlanta is used for the downtown map, while the surrounding interstate system is used for the highway scenario. These road networks provide an environment for the operational model. 1500 agents are added randomly throughout each environment to complete the operational scenario. These agents are provided a speed at the beginning of the scenario using a Gaussian distribution, centered at 25 miles per hour with a standard deviation of 5 miles per hour for downtown driving and 70 miles per hour with 10 miles per hour standard deviation for highway driving. Each agent has a randomly chosen start and end point within the network to create non-uniform behavior. A focus agent was added to capture the operator detection and workload characteristics.

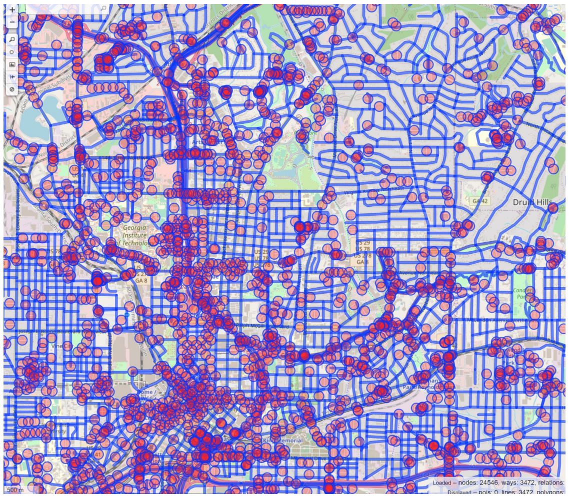

Fig. 7 Data capture from OpenStreetMap used to develop downtown scenario [31].

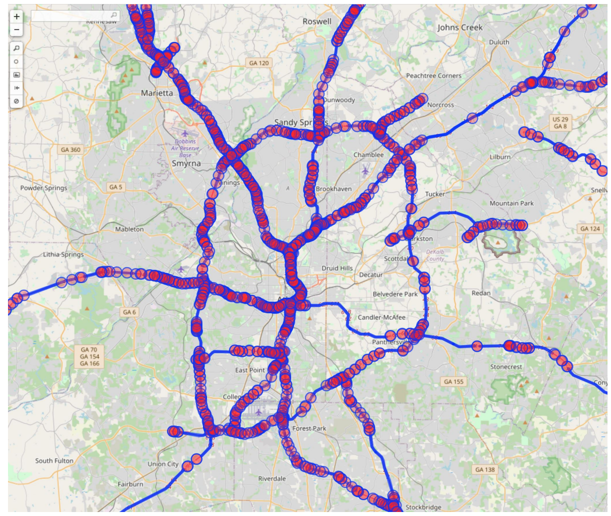

Fig. 8 Data capture from OpenStreetMap used to develop highway scenario [31].

The focus agent has a vision based sensor to detect surrounding vehicles. Based on the direction that the driver is looking, he is capable of seeing a narrow band (+/- 5 degrees) of targets out to 0.1 miles and a wider band (+/- 30 degrees) out to 0.01 miles, representing peripheral detection. If the agent is driving faster than 65 miles per hour (representing divided highways) then it only reports detections heading within 90 degrees of its current heading. A singular agent is chosen to have vision due to run time limitations as the number of interactions increase. Based on the detections created by this vision sensor a driver awareness level is created. It takes 2 seconds for a track, the representation of the other vehicle, to be created and utilized. These tracks are kept in the driver's memory for 6 seconds unless seen again. The task list is currently basic navigation between a start and end point. The focus agent navigates throughout the scenario in the same manner as the other agents, utilizing the road network to calculate the shortest path between a start point and end point.

Figures 9 and 10 show the current implementation. Figure 9 is the downtown model, where cars are driving 25mph, while Figure 10 captures a divided highway operational model, where cars are traveling 70mph. The red car, best seen in Figure 9's top-right corner, is the focus agent and the white agents are the 1500 agents for operational context. These models provide an early workload indicator by tracking the number of car's in the driver's memory. The driver's mental workload is shown for downtown driving in Figure 11 and for highway driving in Figure 12. The blue line in both graphs shows the instantaneous changes in mental demand (number of tracks monitored by the driver). The orange line represents the average number of tracks maintained by the driver. Downtown driving showed a driver with 3.98 tracks on average and a peak of 20, while highway driving had an average of 1.83 tracks and a peak of 11. However, this is only preliminary results for the framework, showing the value of outputting the driver's mental model to see the instantaneous variance and average values.

The preliminary results of this framework demonstrate a dynamic workload metric that can be analyzed along with a visual representation of the environment. Areas of high or low workload can be seen in the visualization to understand the variation in workload. As the model is improved, the resulting workloads from the virtual study can then be compared to experimental data using similar environments and the NASA-TLX scale, such as that from Rahman, Dawal, & Yusoff [29] for downtown, highway, and rural driving. Currently the virtual study differs from the experimental, showing highway driving having half the number of tracks as city driving, whereas in the experimental study [29] the mental loads where similar. This could be due to the road densities being similar (when an interstate could have much higher densities) or the assumption that highway driving should ignore cars traveling in the opposite direction. Further development of the framework and behaviors is required to align virtual and experimental results. Moving to virtual allows for a faster, cost-effective metric that can be used early in the design cycle to assess automation technologies. Utilizing cognitive engineering disciplines to inform the creation of the operations analysis model captures the teaming considerations and the workload metric necessary for system-level trades of automation technology.

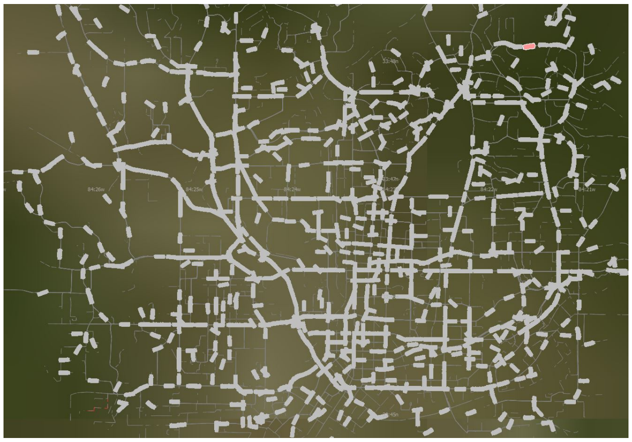

Fig. 9 AFSIM's Results Viewer screenshot showing the city layout and agent density.

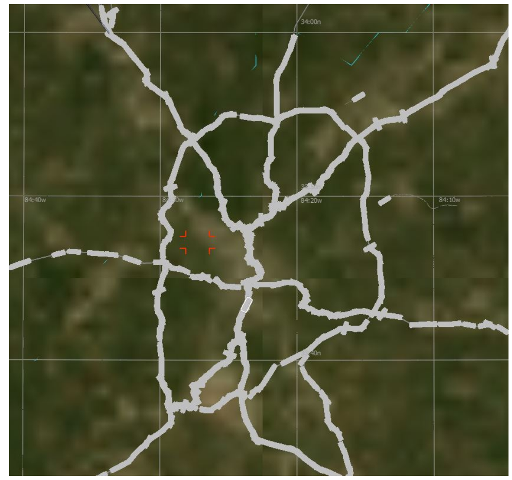

Fig. 10 AFSIM's Results Viewer screenshot showing the city layout and agent density.

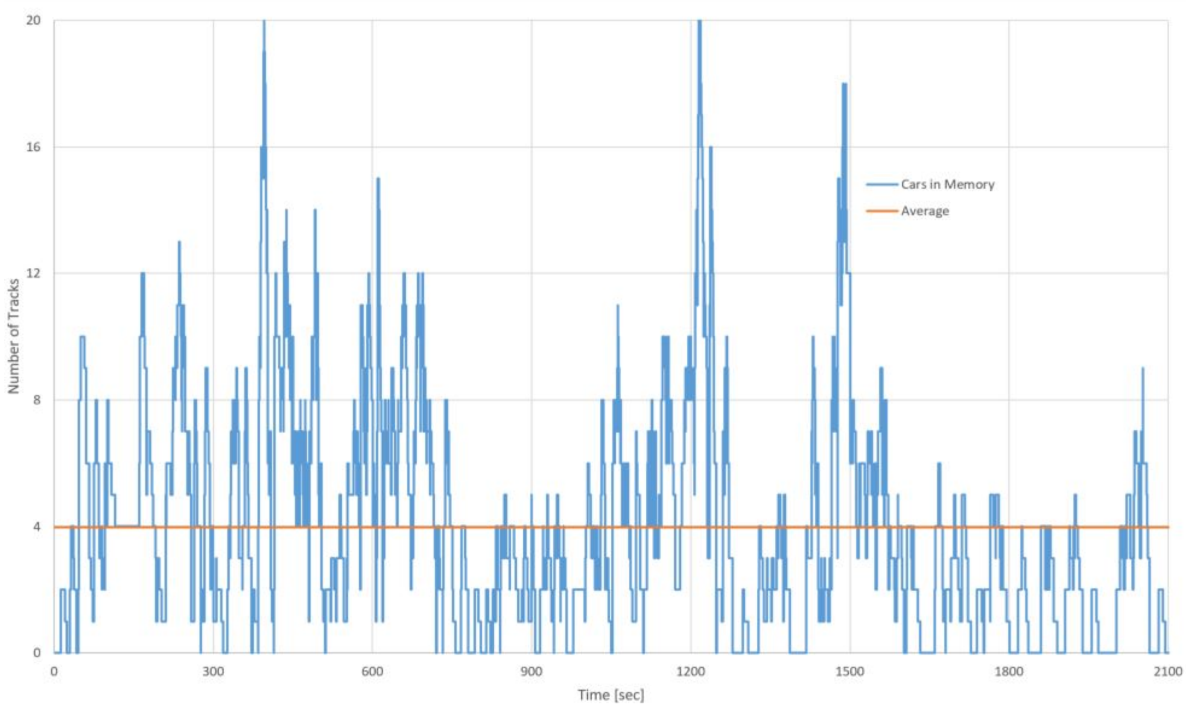

Fig. 11 Driver instantaneous and average mental workload for downtown driving scenario.

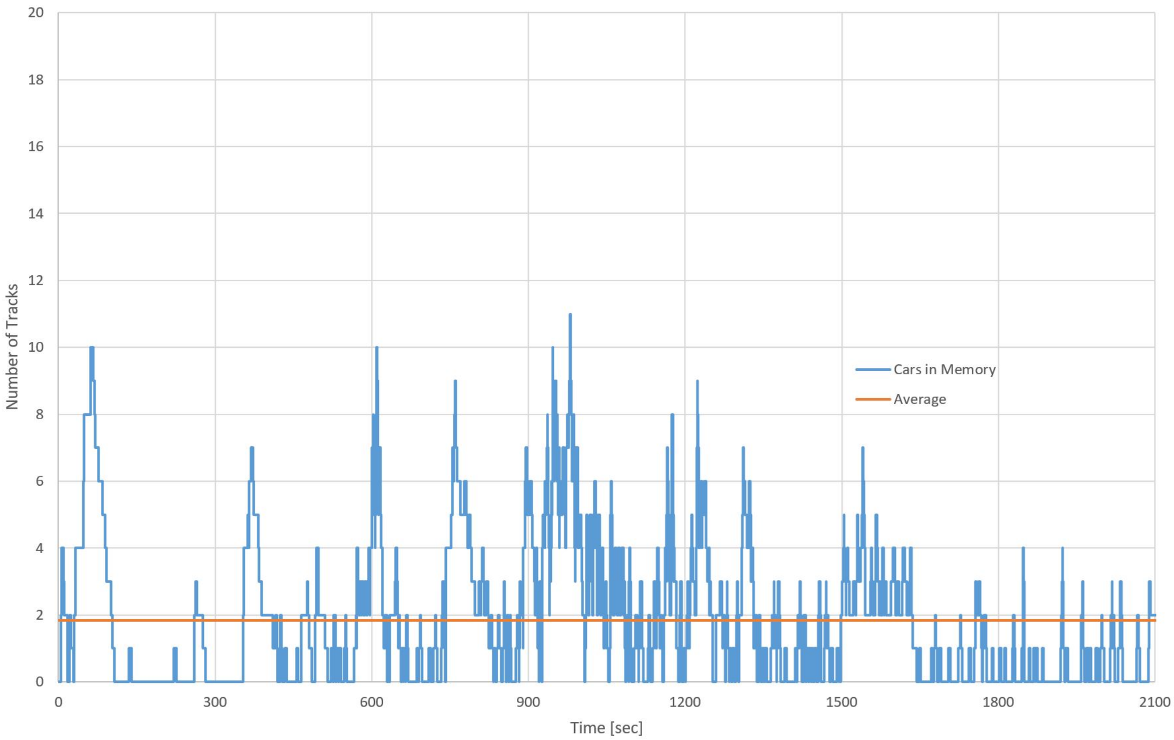

Fig. 12 Driver instantaneous and average mental workload for highway driving scenario.

V. Conclusion

This paper discusses the trend towards system autonomy and the integration of automation along the way. As automation is added, there is a teaming aspect between the automation and the operator that introduces uncertainty during the system design process. This uncertainty requires new methodologies to understand the impacts of automation technology integration during the conceptual design stage. This paper introduces a framework for model creation that expands operations analysis to capture operator workload and awareness through the inclusion of cognitive engineering principles. The reduction in uncertainty and new metrics provided from the model are an enabler for automation technology trades within the conceptual design stage. The preliminary results demonstrated the dynamic measurement component of this larger framework.

This modeling method enables the virtual analyses of automation technologies through explicit tasking to the operator and an operator-focused dynamic workload measurement. Although this is just the first step, the preliminary output shows the value in time dependent workloads. The average mental workload of a driver in the city was only 4 tracks, but the workload spiked up to 20 tracks. This variance and spike demonstrates the risk of overlapping other tasks at the same time as this high mental workload. The next step in framework development is expanding the decision making and workload models of the driver. The current implementation only looks at the mental demands and uses uniform navigation tasks. However, the driver is also conducting safety and route optimization tasks while driving which must be included. These tasks should be mapped together as discussed above using stress based studies to provide a better insight into overall operator workload. As the model is expanded, better alignment should be achieved between the virtual and experimental results. Expanding the behaviors is also required to capture and trade automation technology effects. Technologies impact the decision-action sequence. Since the current model is limited solely to navigation tasks, a limited technology set could be analyzed. The next goal with the driving scenario is to assess SAE level 2 and 3 technologies, like automatic cruise control and lane keeping assistant, therefore further operator task breakdown is required.

This framework reduces the inherent uncertainty in system-level design when not accounting for the teaming aspects between the operator and automation. It provides decision makers necessary insight into the operator's dynamic workload and awareness throughout an operational environment. Studies conducted on UGVs and ATCs, discussed above, highlight the link between operator workload, awareness, and performance, while the failures of autopilot systems by Tesla and the Boeing 737 demonstrate the risks with failing to better understand this link. The same considerations must be taken with future systems such as Boeing's Airpower Teaming System. As automation technologies and autonomy seem to be the way of the future, the necessity for understanding and evaluating these technologies in the conceptual design stage becomes paramount.

References

[1] "Fast-tracking Safe Autonomous Vehicles," *White Paper*, 2018.

[2] "SAE Standards News: J3016 automated-driving graphic update," 2019. URL https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic.

[3] Kaslikowski, A., "Everything you need to know about autonomous vehicles," 2019. URL https://www.digitaltrends.com/cars/the-current-state-of-autonomous-vehicles/.

[4] Coldeway, D., "Boeing's 'Wingman' drone buddies up with pilot-flown jets," 2019. URL https://techcrunch.com/2019/02/26/boeings-wingman-drone-buddies-up-with-pilot-flown-jets/.

[5] Kalra, N., and Paddock, S. M., "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?" *Transportation Research Part A: Policy and Practice*, Vol. 94, 2016, pp. 182–193. URL https://www.rand.org/pubs/research_reports/RR1478.html.

[6] Endsley, M. R., "Design and evaluation for situation awareness enhancement," Vol. 32(2), 1988.

[7] Endsley, M. R., "Toward a Theory of Situation Awareness in Dynamic Systems," *Human Factors*, Vol. 37(1), 1995, pp. 32–64. URL https://www.researchgate.net/profile/Mica_Endsley/publication/210198492_Endsley_MR_Toward_a_Theory_of_Situation_Awareness_in_Dynamic_Systems_Human_Factors_Journal_371_32-64/links/548f61bf0cf214269f263b08.pdf.

[8] Endsley, M., "Situation Awareness Misconceptions and Misunderstandings," *Journal of Cognitive Engineering and Decision Making*, Vol. 9(1), 2015, pp. 4–32. URL https://journals.sagepub.com/doi/pdf/10.1177/1555343415572631.

[9] Klien, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., and Feltovich, P. J., "Ten challenges for making automation a "team player" in joint human-agent activity," *IEEE Intelligent Systems*, Vol. 19(6), 2004, pp. 91–96. URL https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1363742.

[10] Kirschenbaum, S. S., "Uncertainty and Automation," 2003. URL https://apps.dtic.mil/dtic/tr/fulltext/u2/a422305.pdf.

[11] "The Autonomy Paradox," , 2011. URL http://armedforcesjournal.com/the-autonomy-paradox/.

[12] "Driver Errors, Advanced Driver Assistance System Design, Led to Highway Crash," 2019. URL https://www.ntsb.gov/news/press-releases/Pages/NR20190904.aspx.

[13] Woodyard, C., "On autopilot: 'Pilots are losing their basic flying skills,' some fear after Boeing 737 Max crashes," 2019. URL https://www.usatoday.com/story/news/2019/05/25/boeing-737-max-8-autopilot-automation-pilots-skills-flying-hours-safety/1219147001/.

[14] Scovel, C. L., "Perspectives on Overseeing the Safety of the U.S. Air Transportation System," 2019. URL https://www.commerce.senate.gov/services/files/1C1FAD9D-C836-43A6-A54F-839C7C9DD879.

[15] Chen, J. Y. C., and Barnes, M. J., "Robotics Operator Performance in a Military Multi-Tasking Environment," 2007. URL https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6249447.

[16] Chen, J. Y. C., "Concurrent Performance of Military and Robotics Tasks and Effects of Cueing in a Simulated Multi-Tasking Environment," *Presence: Teleoperators and Virtual Environments*, Vol. 18(1), 2009, pp. 1–15. URL https://www.mitpressjournals.org/doi/pdfplus/10.1162/pres.18.1.1.

[17] Edwards, T., Homola, J., Mercer, J., and Claudatos, L., "Multilfactor interactions and the air traffic controller: Ther interation of situation awareness and workload in association with automation," *International Federation of Automatic Control (IFAC)*, 2016, pp. 597–602.

[18] Teigen, K. H., “Yerkes-Dodson: A Law for all Seasons,” *Theory & Psychology*, Vol. 4(4), 1994, pp. 525–547. URL https://www.researchgate.net/profile/Karl_Teigen/publication/247743193_Yerkes-Dodson_A_Law_for_all_Seasons/links/550efd360cf21287416afd07/Yerkes-Dodson-A-Law-for-all-Seasons.pdf.

[19] Kariuki, G., and Lowe, K., “Incorporation of Human Factors in the Design Process,” 2005. URL https://epsc.be/Products/PRISM/_/Background.pdf.

[20] Harris, D., Stanton, N. A., and Starr, A., “Spot the difference: Operational event sequence diagrams as a formal method for work allocation in the development of single-pilot operations for commercial aircraft,” *Ergonomics*, Vol. 58(11), 2015, pp. 1773–1791. URL https://www.tandfonline.com/doi/pdf/10.1080/00140139.2015.1044574?needAccess=true.

[21] Lee, S. E., Olsen, E. C., and Wierwille, W. W., “A Comprehensive Examination of Naturalistic Lane-Changes,” *Ergonomics*, Vol. 58(11), 2015, pp. 1773–1791. URL https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/lane_change_final.pdf.

[22] Hart, S. G., “NASA - Task Load Index (NASA-TLX); 20 Years Later,” 2006. URL https://human-factors.arc.nasa.gov/groups/TLX/downloads/HFES_2006_Paper.pdf.

[23] Hart, S., and Staveland, L., “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research,” *Frontiers in Human Neuroscience*, 1988.

[24] Martin, C., Cegarra1, J., and Averty, P. D., “Analysis of Mental Workload during En-route Air Traffic Control Task Execution Based on Eye-Tracking Technique,” *Engineering Psychology and Cognitive Ergonomics*, 2011, pp. 592–597. URL https://link.springer.com/content/pdf/10.1007%2F978-3-642-21741-8.pdf.

[25] Tolk, A., “Merging Two Worlds: Agent-Based Simulation Methods for Autonomous Systems,” *Modeling, Simulation & Visualization Engineering Faculty Publications*, 2015. URL https://digitalcommons.odu.edu/msve_fac_pubs/32.

[26] Hester, P. T., and Tolk, A., “Applying Methods of the M&S Spectrum for Complex Systems Engineering,” 2010. URL https://dl.acm.org/doi/pdf/10.1145/1878537.1878615?download=true.

[27] West, T. D., and Birkmire, B., “AFSIM: The Air Force Research Laboratory’s Approach to Making M&S Ubiquitous in the Weapon System Concept Development Process,” *Journal of Cyber Security and Information Systems*, Vol. 7, No. 3, 2020. URL https://www.csiac.org/journal-article/afsim-the-air-force-research-laboratorys-approach-to-making-ms-ubiquitous-in-the-weapon-system-concept-development-process/.

[28] Clive, P. D., Johnson, J. A., Moss, M. J., Zeh, J. M., Birkmire, B. M., and Hodson, D. D., “Advanced Framework for Simulation, Integration and Modeling (AFSIM),” *International Conference Scientific Computing*, 2015. URL http://worldcomp-proceedings.com/proc/p2015/CSC7058.pdf.

[29] Rahman, N. I. A., Dawal, S. Z. M., and Yusoff, N., “Subjective responses of mental workload during real time driving: A pilot field study,” *IOP Conference Series: Materials Science and Engineering*, Vol. 210, 2017. URL https://iopscience.iop.org/article/10.1088/1757-899X/210/1/012076/pdf.

[30] Schneider, M., McGrogan, J., Colombi, J. M., Miller, M. E., and Long, D. S., “Modeling Pilot Workload for Multi-Aircraft Control of an Unmanned Aircraft System,” *INCOSE International Symposium*, Vol. 21(1), 2011, pp. 796–810.

[31] “OpenStreetMap,” 2020. URL https://www.openstreetmap.org/copyright.